(2)

INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT

AD-A206 803

Unisys/Defense Systems

ARPA ORDER NUMBER: 5262
PROGRAM CODE NO. NR 049-602 dated 10 August 1984 (433)
CONTRACTOR: Unisys Defense Systems
CONTRACT AMOUNT: $1,704,901
CONTRACT NO: N00014-85-C-0012
EFFECTIVE DATE OF CONTRACT: 4/29/85
EXPIRATION DATE OF CONTRACT: 4/28/89
PRINCIPAL INVESTIGATOR: Dr. Lynette Hirschman
PHONE NO. (215) 648-7554

SHORT TITLE OF WORK: DARPA Natural Language Understanding Program

REPORTING PERIOD: - 11/1/88 - 1/31/89

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

DTIC
ELECTE
APR 1 2 1989
S D
H

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

89 3 23 045

1. Description of Progress

The highlights of this three month period were:

- The Natural Language Evaluation Workshop, organised and hosted by Unisys;
- The DARPA February Speech and Natural Language Workshop,
  - * organising the workshop;
  - * porting PUNDIT to a new domain and reporting the results
- Progress in the development of the semantics-based selection mechanism;
- Response to the DARPA BAA for a new joint effort in message understanding;
- A Report on a Performance Task for Written Language Understanding;
- Preparation for the MUCK-II Message Understanding Conference;
- Progress in related areas not funded by DARPA:
  - * Interfacing PUNDIT to the MIT Speech Recognition System;
  - * A Performance Task for Spoken Language
  - * Progress on integrating the NLM Lexicon and the PUNDIT system.

1.1. Evaluation of NL systems

A workshop on evaluation on natural language processing systems was held at the Wayne Hotel in Wayne, Philadelphia, Dec 8,9. There were 50 participants from the US, Europe and Canada, including university professors, representatives of funding agencies, and people from industry and government. The workshop was organised by Martha Palmer (Unisys) with assistance from a committee comprised of Lyn Bates (BBN), Beth Sundheim (NOSC), Tim Finin (Unisys), Mitch Marcus (U Penn) and Ed Hovy (ISI). The workshop discussed evaluation methods used in other disciplines and examples of attempts to evaluate natural language processing systems. The participants then broke up into working groups to discuss how to apply those methods to particular areas in natural language processing. The working groups came up with specific proposals for evaluation methods and for additional workshops focused on specific applications. There was a panel discussion on the need for a large corpus of both written and spoken language that could be used for training sets and test sets for specific applications. Mitch Marcus (U Penn) has a proposal in to Darpa to build such a corpus. There is general agreement that syntactic parsers at least are ready for some type of systematic evaluation and comparison based on training sets and test sets. Areas such as semantics and pragmatics would benefit from continued discussion of appropriate types of evaluation, and a move towards a consensus on representations. With respect to specific applications rather than components, a workshop on evaluating message understanding systems will be held next June to compare several different systems on portability and performance based on a sample of 100 messages. Beth Sundheim (NOSC) is organising it. Similar workshops to compare question-answering systems and generation systems were proposed.

1.2. Preparation for DARPA Speech/Natural Language Meeting

Our major focus during January has been prepartion for the DARPA Speech and Natural Language meeting, to be held in February in Philadelphia. Unisys is not only a participant, but Lynette Hirschman is the General Chair for the Workshop, responsible for the overall planning as well as local arrangements. The focus of the meeting will be to emphasise Spoken Language Systems, and to encourage communication between the Speech researchers and the Natural Language researchers.

1.2.1. Unisys Presentations at the DARPA Workshop

In its technical role, Unisys will be making a total of four presentations:

1. Report on the Natural Language Evaluation Workshop;
2. Report on the Spoken Language Research at Unisys;
3. Report on the Natural Language Processing Research;
4. Report on an Automated Maintenance Assistant as a Performance Task for Spoken Language.

In addition to these presentations, we plan to submit three papers for publication in the Conference Proceedings. These are listed at the end of the Report.

1.2.2. Port to the Resource Management Domain

A major focus of recent work has been porting PUNDIT to a new domain, the Resource Management domain, which consists of queries to a database about ship movements and characteristics. Our goal is to report figures for porting the syntactic and regularisation components of the system at the DARPA meeting. This is the first sisable non-message domain that we have ported PUNDIT to, and we expect that a successful port will demonstrate the generality of the PUNDIT system and our tools for bringing up PUNDIT in a new application domain.

1.3. Syntax/Semantics Interaction

We have been debugging the new semantics-based selection mechanism, which integrated syntax and semantics. This mechanism is based on the semantic interpreter, and it now successfully handles all of the verb phrases and nominalisations in CASREPS. We have started porting it to MUCK, and have discovered the necessity of selection data base patterns for certain areas such as noun-noun compounds and adverbs which are not handled by the semantic interpreter. We are looking into the possibility of merging the semantic intepreter parse filtering with SPQR.

1.4. Response to DARPA BAA

We are writing a collaborative proposal with SRI and NYU in response to the DARPA BAA for intelligence message analysis. The proposal is aimed at extending PUNDIT to include a more sophisticated knowledge representation component that will perform the types of reasoning required by the MUCK II application. This will permit the integration of our tool for adding lexical items, Knacq0, funded internally (see Reston below), with a tool for building and maintaining a domain model. This integration will allow all of the relevant information about a lexical item to be added at one time. As a prerequisite for this integration we will be extending our notion of a verb taxonomy - the representation of the relationships indicated by the verbs in the domain model - to more closely tie together the verb semantics and the domain model.

1.5. Performance Task for Written Language Understanding

We have defined a black box information retrieval task for PUNDIT to be used to provide a quantitative evaluation of PUNDIT's performance. This task involves inputing messages to the system for analysis, storing the results of the analysis in a database, and retrieving information via natural language questions to the system. The task has been very satisfactory because it is easy to perform and evaluation of the results is completely mechanical. In contrast, other proposed evaluation tasks for natural language have often been very labor intensive to implement and require a high level of expertise in order to evaluate the results.

In this task, we compared PUNDIT performance to that of keyword-based retrieval and that of human retrievers on four questions for both test and training data. The results are shown below for 17 training and 17 test messages from the RAINFORM domain. Measures of both recall (percent of relevant messages retrieved) and precision (percent of retrieved messages relevant) are given.

Results: TRAINING

Processor	Recall	Precision
Human	98%	100%
PUNDIT	24%	100%
Keyword	29%	86%

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By per letter
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

Results: TEST

Processor	Recall	Precision
Human	100%	81%
PUNDIT	13%	100%
Keyword	53%	66%

An analysis of the data indicated that the major reason for recall failure was PUNDIT's inability to draw inferences. For example, if the message stated, *fired asroc and torpedos,* PUNDIT is not currently capable of recognising this situation as an attack. We plan to remedy this with the work described above in Section (MUCK II).

1.6. Preparation for MUCK II

In preparation for the upcoming MUCK II conference on message understanding, we have begun to define the requirements for the database update task. Analysis of the database formats proposed by Beth Sundheim at NOSC indicate that significant inferential capabilities will be required in order to complete the database update task correctly. PUNDIT currently produces a very literal interpretation of the input and does not attempt to make extensive inferences. This design decision has been effective in allowing us to focus on the processing of the literal meaning of inputs, but now needs to be augmented by additional, more general, reasoning capabilities. We have defined two tasks to address this need.

First, in order to provide a well-defined basis for inference we are re-implementing the knowledge bases in M-PACK, which is a semantic net based knowledge representation system developed at Unisys, similar to KL-ONE in design. A tool for converting formatted ASCII text specifications of a knowledge base into an M-PACK format has been developed with approximately 2 person days of effort. We estimate that this tool gave us about a 2-fold increase in efficiency in initial entry of the knowledge base. However, the main time savings from this tool is estimated to be in the area of correcting and revising the model. Previously, in the worst case, certain revisions required re-entry of the entire knowledge base, a several day task. Now, revisions can be made in a few minutes by simply editing an ascii file.

The second major task for MUCK II is to add inferential capabilities to system. We are currently defining requirements for this task.

2. Description of Related Progress Under Other Contracts

We describe below some key developments in the overall PUNDIT system that have been funded by other sources (Unisys IR&D, NLM contract), since these contribute significantly to the overall development of the system.

2.1. Portability

In order to facilitate the technology transfer of PUNDIT to Reston we focused on improving our tools for adding lexical items. We greatly improved the ease of use and help facility of the Lexical Entry Procedure, and we also implemented from scratch a Semantic Rule Editor for inputing and editing information about verb semantics. These are now linked together so that the lexical entry procedure can be called while editing the verb semantics. We presented a group at Unisys Defense Systems in Reston with a proposal for follow-on work to improve the link between the LEP and the SRE and to provide for more guidance in the selection of thematic roles.

2.2. Spoken Language IR&D

We have successfully demonstrated a methodology for interfacing PUNDIT to the MIT speech recognition system. This involves the use of word lattice or network, and a technique for traversing this network. The network traversal technique partitions the network, which allows the system to do a "best-first" exploration of possible word sequences in a reasonable amount of time. This work is described in the paper by John Dowding, "Reducing Search by Partitioning the Work Network". We are also running experiments to apply PUNDIT to the output of the word

network traversal, to see how this improves word accuracy. We plan to present these measures of word accuracy, as well as figures on perplexity, at the DARPA Workshop in February.

2.3. Performance Task for Spoken Language

In preparation for work in Spoken Language Systems, we have been exploring the feasibility of developing a spoken language interface to an expert system as a performance task. The particular expert system we are looking at is a maintenance assistant for an optical character reader (OCR), which is being developed at the Paoli Research Center for the US Postal Service. The expert system, KSTAMP, is attractive as a useful application for spoken language processing for the following reasons:

(1) It is already a real application doing a real job for a real customer.

(2) For speech, it offers a situation where speech input would be a definite advantage. That is, as things are now, the maintenance person has his or her hands inside the OCR machine (which is large), but to receive instructions, the person has to stop work, walk to the computer screen, and type input to KSTAMP. Therefore, a headset microphone allowing the operator to speak input to KSTAMP would save time and annoyance.

(3) For natural language, KSTAMP also offers an opportunity to make the process considerably more efficient. As things are now, the operator is led through a series of menu choices to an identification of the specific maintenance problem. This process is time-consuming. If the operator were speaking to a person instead of a machine, s/he could describe the problem much more efficiently in one natural English sentence.

(4) A KSTAMP application allows for clear definition of performance evaluation tasks, an opportunity to run experiments, and ability to measure the results. Testing of NL and speech could proceed independently. Among the things that could be measured are: time saved over the original system by use of speech input, time saved by allowing the operator some freedom in the phrasing of input, time saved by allowing the operator to by-pass the menu system, and time saved in training.

The KSTAMP system currently employs a vocabulary of 399 domain-specific words. A fully developed spoken language system would probably require several hundred additional domain-independent words such as pronouns, prepositions, and other general-purpose vocabulary.

We have defined three stages in the development of this performance task.

Stage 1: Replace typing input with speech recognition. It would also be useful to have a speech synthesis capability so that the operator doesn't have to leave work to look at KSTAMP's replies on the terminal.

Stage 2: Incorporate simple natural language processing tasks as in, for example, allowing the operator to say either *excessive verifier alarms, verifier alarms excessive,* or *too many verifier alarms.*

Stage 3: Use continuous speech recognition and whole-sentence natural language input to identify repair problems.

2.4. Progress under the NLM Contract

In December we completed the first six months of work on the contract for Automated Analysis of Biomedical Text. The ultimate objective of building such a tool is to test the hypothesis that access to a bibliographic database, such as MEDLINE, can be improved by automated analysis of the free text found in the title and abstract fields of MEDLINE citation records. The major focus of work to date has been on integrating the NLM lexicon with the syntactic processing components of UNISYS' PUNDIT system. This has given us experience in converting the large-scale NLM lexicon into a form suitable for PUNDIT, and has also raised questions about processing complex technical prose (in distinction to military messages, our main application areas to date).

3. Change in Key Personnel

None.

4. Summary of Substantive Information from Meetings and Conferences

4.1. Darpa Meetings

(1) Meeting between Lynette Hirschman and Charles Wayne, to plan DARPA Speech and Natural Language Workshop, Jan. 6, 1989.

4.2. Papers and Presentations

Catherine Ball, "Analysing Explicitly Structured Discourse in a Limited Domain: Trouble and Failure Reports", to appear in the *Proc. of the Workshop on Speech and Natural Language*, sponsored by DARPA ISTO.

John Dowding, "Reducing Search by Partitioning the Word Network" accepted for presentation at the AAAI Symposium on Spoken Language Systems, Stanford, March 28-30, 1989; will also appear in the *Proc. of the Workshop on Speech and Natural Language*, sponsored by DARPA ISTO.

Lynette Hirschman, Francois-Michel Lang, John Dowding, Carl Weir, "Porting PUNDIT to the Resource Management Domain", to appear in the Proc. of the Workshop on Speech and Natural Language, sponsored by DARPA ISTO.

R. Passonneau, "Getting at Discourse Referents", submitted to the 1989 Annual Meeting of the Association for Computational Linguistics, Vancouver, June, 1989.

4.3. Conference Attendance

Martha Palmer, Deborah Dahl, Tim Finin, and Lynette Hirschman attended the Natural Language Evaluation Workshop, held in Philadelphia, Dec. 8-9.

5. Problems Expected or Anticipated

Unisys has submitted paperwork requesting a no-cost extension through October 31, 1989. This has been necessitated by the delay in receipt of 1988 funding.

6. Action Required by the Government

Approval of the requested no-cost extension through September, 1989 is needed. Paperwork on the final $13K increment is at ONR and should be received shortly.

7. Fiscal Status

(1) Amount currently provided on contract:
$ 1,691,157 (committed funding) $ 1,704,901 (contract value)

(2) Expenditures and commitments to date:
$ 1,355,736

(3) Funds required to complete work:
$ 349,165